International Journal of Medical and Pharmaceutical Research

Online ISSN-2958-3683 | Print ISSN-2958-3675
Frequency: Bi-Monthly
Available online on: https://ijmpr.in/

Original Article

Enhancing Efficiency in MCQ Evaluation: Integrating Voice-to-Text and Excel Automation in Medical Education

Dr. Sankalp[1], Dr. Hanjabam Barun Sharma[2], Dr. Thakur Nidhi*[3], Dr. Ashish Kumar Gupta[13]

[1]Assistant Professor, Dept. of Physiology, Sports and exercise laboratory, 2nd floor, Dept of Physiology, Institute of Medical Sciences, Banaras Hindu University, Aurobindo Colony, Bhelupur, Varanasi, UP 221005
[2]Associate Professor, Physiology, Sports and exercise laboratory, 2nd floor, Dept of Physiology, Institute of Medical Sciences, Banaras Hindu University, Aurobindo Colony, Bhelupur, Varanasi, UP 221005
[3]Assistant Professor, Biochemistry, ESIC Medical College and Hospital, Varanasi (U.P), India
[4]Assistant Professor, Physiology, Institute of Medical Sciences, Banaras Hindu University, Lanka, Varanasi, UP 221005

OPEN ACCESS

Corresponding Author:

Dr. Thakur Nidhi
Assistant Professor, Biochemistry, ESIC Medical College and Hospital, Varanasi (U.P), India

Received: 14-10-2025
Accepted: 29-10-2025
Available online: 12-11-2025

ABSTRACT

Background: Multiple-choice questions (MCQs) are widely used in medical education due to their efficiency in assessing a broad range of knowledge. However, traditional grading of MCQs can be labor-intensive and prone to human error, especially in large volumes. With increasing class sizes, there is a growing need for more efficient and accurate grading systems.
Objective: This study evaluates the effectiveness of integrating voice-to-text technology (VTT) and Excel automation to enhance the efficiency and accuracy of MCQ grading in medical education.
Methods: A total of 3,000 simulated MCQs responses were evaluated using both manual and automated methods. VTT technology was used to transcribe responses, which were then organized into a table using a Large Language Model (LLM) for data parsing and structuring and scored in Excel using automation techniques. Two evaluators assessed the papers, and time taken, error rates, evaluator fatigue, and satisfaction were compared between methods.
Results: The automated method significantly reduced evaluation time to 45.9 minutes (0.76 man-hours) compared to 194.82 minutes (3.25 man-hours) for manual evaluation, with a similar high accuracy rate of 99.96% for both methods. Evaluator fatigue was lower, and satisfaction was higher with the automated method.
Conclusion: The integration of VTT technology and Excel automation significantly improves the efficiency of MCQ grading while maintaining high accuracy. This approach offers a scalable, cost-effective solution for medical education settings, particularly in resource-limited environments. Future research could expand this method to other types of assessments and explore additional automation tools to further enhance educational evaluation processes.

Keywords*: MCQ Evaluation; Voice-to-Text Technology; Excel Automation; Medical Education Assessment; Grading Efficiency*

INTRODUCTION

Multiple-choice questions (MCQs) are a fundamental component of assessment in medical education, widely used for their ability to efficiently test a broad range of knowledge (Haladyna, Downing, & Rodriguez, 2002). Despite their popularity, grading MCQs, especially in large volumes, can be labor-intensive and susceptible to human error, potentially impacting the reliability and validity of assessments (Schuwirth & van der Vleuten, 2011). With increasing class sizes and the need for frequent assessments, there is a growing demand for more efficient and accurate grading systems in medical education (Case & Swanson, 2002).

Technological advancements provide new opportunities to address these challenges. Tools like voice-to-text software and Excel automation have shown promise in reducing the workload associated with grading while maintaining, if not enhancing, accuracy (Peat, 2000). Voice-to-text technology offers a hands-free way to enter data, which, when combined with Excel's powerful automation capabilities, can significantly streamline the grading process (Tobin, 1998). This study explores the effectiveness of integrating voice-to-text and Excel automation in the evaluation of MCQs in a medical education setting. The goal is to compare the efficiency, accuracy, and evaluator satisfaction between traditional manual grading and a technologically enhanced automated approach.

Methods

MCQ Evaluation Process

The study evaluated a total of 3,000 MCQs simulating the theory assessments of first-year MBBS students in Physiology. The MCQs were distributed across 300 exam sheets, with each sheet containing 10 MCQs, each having four options (A, B, C, D). These exam sheets were further divided into 15 bundles, each comprising 20 papers. Each bundle was timed separately to assess the processing efficiency of the different evaluation methods—Voice-to-Text (VTT) and manual.

Step	Voice-to-Text (VTT) Method	Manual Method
1. Collection of MCQ Sheets	Collect and organize MCQ sheets into bundles	Collect and organize MCQ sheets into bundles
2. Transcription	Transcribe responses using Google Docs' voice-to-text	Read and mark responses directly on MCQ sheets
3. Error Checking	Use keyboard shortcuts in Google Docs for quick corrections during transcription	Continuous cross-checking and manual corrections on paper
4. Conversion to Table	Convert text responses to a structured table using ChatGPT	Not applicable
5. Data Entry into Excel	Transfer formatted data from ChatGPT to Excel	Manually enter scores and data into Excel
6. Automated Scoring	Apply Excel formulas for automated scoring	Not applicable; manual scoring on the sheets
7. Final Consistency Check	Verify responses and automated scores against answer key	Verify total marks and ensure they match the manual calculations
8. Final Scoring	Automated calculation of total marks using Excel	Manually calculate total marks after entering data into Excel

Figure 1 : Workflow Diagram of MCQ Evaluation Methods

As shown in **Figure 1**, During the VTT method, the evaluator transcribed each MCQ's response by verbally indicating the options as "Alpha," "Bravo," "Charlie," and "Delta" to avoid confusion during the voice-to-text conversion process. The serial numbers of the MCQ papers were read sequentially, followed by the chosen options for each question. If a question was left unanswered or had multiple answers, the term "Fun" was spoken. The VTT method was conducted first for all 300 sheets to transcribe responses and calculate scores using automated processes. After completing the VTT evaluation for all sheets, the manual method was then employed. During the manual evaluation, the total score for each sheet was recorded directly on the sheet after scoring all 10 MCQs, and this score was included in the processing time for each manual evaluation. Two evaluators participated in the assessments: Evaluator 1 processed seven bundles, while Evaluator 2 handled the remaining eight bundles.

Technology and Tools Used

The recorded time for each method represented the duration required to process the MCQ sheets per bundle. For the VTT method, processing time included transcribing the responses using Google Docs' voice-to-text feature, set to "English (United Kingdom)" with the "English (India)" variant to match the evaluators' accents. The in-built laptop microphone was used with optimized settings for accurate input. Dictation was conducted at a moderate pace to capture each word correctly, as speaking too quickly often led to incomplete or inaccurate responses. Brief pauses were made between dictating answers to allow the software time to process the input and reduce errors. A stopwatch feature on an Android phone was used to

accurately record the timing for each evaluation task. The processing time recorded for the VTT method did not include the calculation of marks, which required subsequent steps in Excel after organizing the data into a table using a Large Language Model (LLM) for data parsing and structuring.

The manual method's processing time included immediate scoring of the MCQ sheets, but similar to VTT, an additional 25 minutes was required to manually enter these scores into Excel, as shown in **Table 1**. This time for data transfer was excluded from the overall time calculation for both methods.

Table 1: Comparison of Steps in Voice-to-Text (VTT) and Manual MCQ Evaluation Methods

Step	VTT Method	Manual Method
1. Collection of MCQ Sheets	Collect and organize MCQ sheets into bundles	**Collect and organize MCQ sheets into bundles**
2. Transcription	Use Google Docs' voice-to-text to transcribe responses	**Directly read and mark responses on MCQ sheets**
3. Error Checking	Use keyboard shortcuts for corrections during transcription	**Continuous cross-checking during manual marking**
4. Data Formatting	Convert text to Excel format using LLM	**Not applicable**
5. Scoring	Automated scoring using Excel formulas	Manual scoring with immediate total calculation
6. Data Entry into Excel	Copy formatted data into Excel	Manually enter scores into Excel
7. Final Consistency Check	Verify entries against answer key; review automated scores	Verify total marks match manual calculations
8. Time Required per Bundle	Time recorded for VTT processing (excluding Excel transfer time)	Time recorded for manual marking and entry (excluding Excel transfer time)

Evaluator Fatigue and Satisfaction Scores

To assess the impact of each evaluation method on evaluator well-being and preference, fatigue and satisfaction scores were recorded after the completion of all bundles assigned to each evaluator. Fatigue was measured on a scale of 1 to 5, with 1 indicating no fatigue and 5 indicating extreme fatigue. Satisfaction was similarly rated from 1 to 5, with 1 representing low satisfaction and 5 representing high satisfaction. Each evaluator provided a fatigue and satisfaction score twice—once after completing all the bundles using the VTT method and again after completing all the bundles using the manual method. These scores were collected to evaluate differences in fatigue and satisfaction between the two methods, as shown in **Table 2**.

Table 2: Time, Accuracy, and Evaluator Experience Metrics for Voice-to-Text (VTT) and Manual Methods

Metric	VTT Method	Manual Method
Total Time for 3000 MCQs (minutes)	45.9 (0.76 man-hours)	194.82 (3.25 man-hours)
Average Time per Sheet (seconds)	9.18	38.96
Total Errors	12	13
Error Rate	0.04	0.043
Accuracy (%)	99.96	99.96
Evaluator Fatigue Score	3	4
Satisfaction Score	5	2

Command Used for LLM
The following command was used to convert the Voice-to-Text data into a table format suitable for Excel: "Please convert the following Voice-to-Text data into a table format suitable for Excel. The data includes the serial number of each MCQ paper followed by the responses to ten questions. Each response is indicated by a keyword: 'Alpha' for option A, 'Bravo' for option B, 'Charlie' for option C, 'Delta' for option D, and 'Fun' for any unanswered or multiple-marked questions. Organize this data into a table with the following column headers: 'Serial No', 'Q1 Response', 'Q2 Response', 'Q3 Response', 'Q4 Response', 'Q5 Response', 'Q6 Response', 'Q7 Response', 'Q8 Response', 'Q9 Response', 'Q10 Response'. Ensure each row starts with the serial number and includes the corresponding responses in separate columns."

Pre-Processing Procedures
Calibration of Voice Typing: Short practice sessions were conducted with each evaluator using the dictation feature to help them become accustomed to the system and to allow the software to adapt to their unique voice patterns.

Noise Reduction Measures: Dictation was performed in a quiet room to minimize background noise and potential distractions. All other background applications on the computer were closed to reduce system noise and enhance the performance of Google Docs' voice-to-text feature. These steps were taken to ensure optimal transcription accuracy.

Error-Checking Procedures
Real-Time Monitoring: Real-time monitoring was employed during both VTT and manual methods to ensure accuracy. For the VTT method, evaluators used keyboard shortcuts for quick corrections during dictation (e.g., Ctrl + Z to undo, Ctrl + Delete to delete the last word, Enter to separate new entries). Automated error detection features, such as spell check and grammar tools in Google Docs, were disabled to prevent unnecessary flagging of non-standard entries as errors. In the manual method, evaluators cross-checked their work continuously to minimize human errors due to oversight or fatigue.

Post-Dictation Review: After processing each bundle of MCQ sheets, additional error-checking was performed. In the VTT method, the transcribed responses were intermittently compared against the original answer sheet or key to verify accuracy and make necessary adjustments before organizing the data into a table using LLM for Excel scoring. For the manual method, evaluators reviewed each marked sheet before entering their respective final scores into Excel, ensuring that the marks recorded were accurate reflections of the answers provided.

Consistency Check: As depicted in **Figure 2**, custom validation rules were applied in Excel to identify entries that did not match the predefined answer options (A, B, C, D), highlighting potential errors before final scoring. For visual verification, cells containing "Alpha" were marked red, "Bravo" as blue, "Charlie" as green, "Delta" as yellow, and "Fun" as black, making it easier to spot any incorrect or missing entries. In the VTT method, both the responses recorded and the marks awarded could be cross-verified against the original MCQ sheets to ensure accuracy. In contrast, for the manual method, only the total marks awarded could be compared to ensure they matched the calculated totals, without verifying each individual response. This difference allowed a more detailed consistency check for the VTT method than for the manual method.

Serial No	Q1 Response	Q2 Response	Q3 Response	Q4 Response	Q5 Response
1	Alpha	Bravo	Charlie	Delta	Fun
2	Bravo	Charlie	Delta	Alpha	Bravo
3	Charlie	Delta	Alpha	Bravo	Fun
4	Delta	Alpha	Bravo	Charlie	Alpha
5	Fun	Bravo	Charlie	Delta	Bravo

Figure 2: Example of Error-Checking Setup in Excel

Automated Scoring System
After copying the data to the Excel sheet, the responses were converted from "Alpha," "Bravo," "Charlie," and "Delta" back to "A," "B," "C," and "D." An array formula was used to match each response against the answer key and award marks accordingly. The second row had column headers with serial numbers, followed by Q1 to Q10 response columns. Adjacent columns were designated for marks awarded for each question and a total marks column. The array formula to match responses was applied for all rows, and a sum function was used to calculate the total marks for each MCQ paper.

Evaluator Training and Calibration

No formal training was provided, as the evaluators were already familiar with Google Docs and Excel. However, a short practice session was conducted to familiarize them with the voice-to-text process. No calibration was conducted, but evaluators were instructed to evaluate papers manually as quickly as possible, taking breaks between bundles if needed.

Statistical Analysis

A paired t-test was conducted to compare the time taken for MCQ evaluation between the Voice-to-Text (VTT) and Manual methods, resulting in a p-value close to zero, confirming a highly significant difference favoring the VTT method. Correlation analysis revealed a weak negative correlation for the Manual method (-0.2107) and a moderate negative correlation for the VTT method (-0.5391), indicating that time investment in VTT may reduce errors more effectively, as detailed in **Table 2**.

Results

The study evaluated 3,000 multiple-choice questions (MCQs) using both manual and automated methods. The automated method (Voice-to-Text, VTT) significantly reduced evaluation time to 45.9 minutes (0.76 man-hours) compared to 194.82 minutes (3.25 man-hours) for manual evaluation, resulting in an average time per paper of 9.18 seconds for VTT versus 38.96 seconds for the manual method. **Figure 3** illustrates the substantial time savings achieved when using the VTT method for MCQ evaluation compared to the manual method, clearly showing the efficiency advantage of the VTT approach.

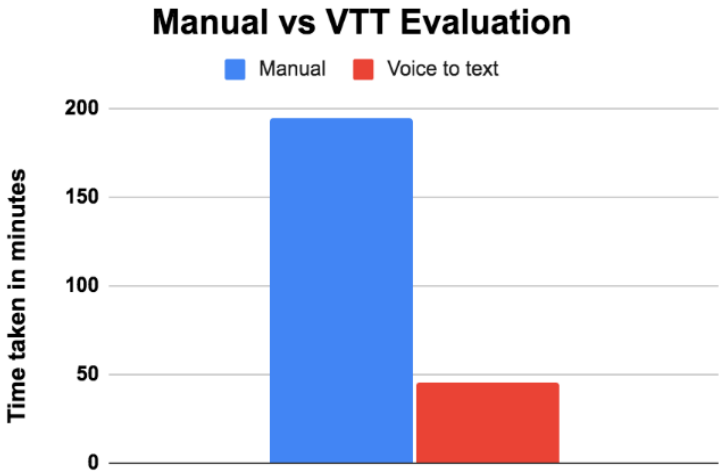

Figure 3. Time Taken for MCQ Evaluation Using Manual vs. Voice-to-Text Methods

In terms of errors, the manual method resulted in 13 errors, leading to an error rate of 0.043 per paper and an accuracy of 99.96%, while the automated method had 12 errors, resulting in an error rate of 0.04 per paper and an accuracy of 99.96%. The distribution and contribution of errors from both methods to the total number of errors observed are depicted in **Figure 4**, which highlights how each method contributed to the total errors rather than displaying the error rates in isolation.

Figure 4: Contribution of Error Distribution opportunities in MCQ Evaluation: Voice-to-Text & Manual Methods

Figure 5 provides a detailed view of the average time taken per MCQ paper across different bundles, demonstrating that the VTT method consistently outperformed the manual method in terms of time efficiency per paper. This line chart shows a clear trend favoring the automated method across all bundles evaluated.

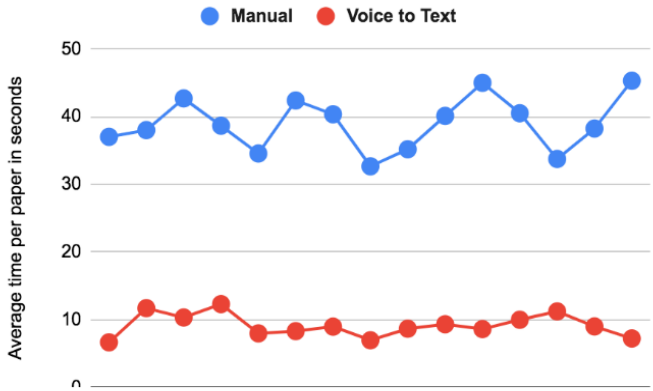

Figure 5: Average Time of Processing Per Paper: Manual vs. Voice-to-Text

Further analysis using scatter plots revealed insights into the relationship between time taken and errors in both methods. **Figure 6** displays the correlation between time taken and the number of errors using the VTT method, showing that the number of errors does not significantly increase with the time taken, suggesting the reliability of the VTT method over varying time spans.

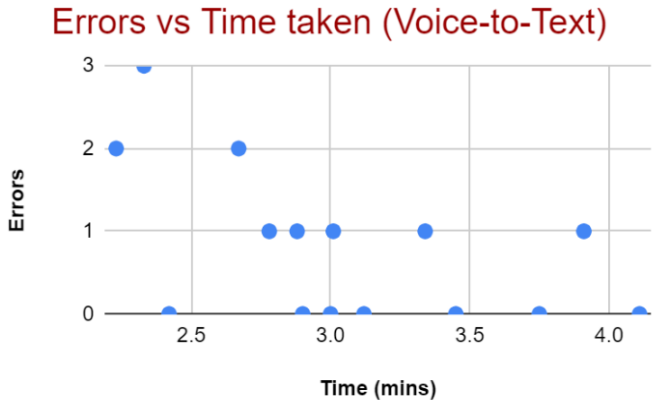

Figure 6: Errors vs. Time Taken Using Voice-to-Text Method

In contrast, **Figure 7** illustrates the errors versus time taken in the manual method, indicating a potential increase in errors as more time is spent on evaluation, possibly due to evaluator fatigue or inconsistencies.

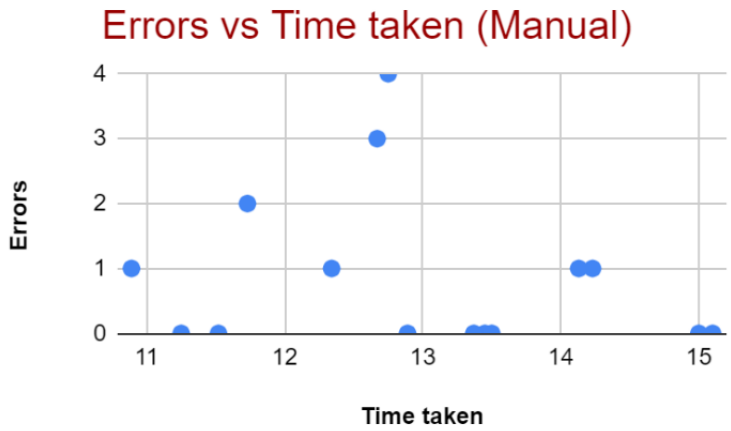

Figure 7: Errors vs. Time Taken Using Manual Method

Evaluator fatigue and satisfaction were also assessed for both methods. Evaluator fatigue was lower with the VTT method, scoring 3 out of 5, compared to 4 for the manual method, and evaluator satisfaction was higher with the VTT method, scoring 5 out of 5, versus 2 for the manual method. These scores are detailed in **Table 2**, which provides a comprehensive comparison of time, accuracy, and evaluator experience metrics for both methods.

Table 3 further breaks down the time, errors, and performance metrics across 15 bundles of 200 MCQs each, comparing the VTT and manual evaluation methods. This table gives a comprehensive overview of the time taken per bundle, error rates, average time per MCQ paper, and evaluator distribution, allowing for a detailed comparison of both methods.

Table 3: Comparison of Time, Errors, and Evaluator Performance in MCQ Evaluation Methods

Bundle #	MCQs per Bundle	Time (mins) - Voice-to-Text	Time (mins) - Manual	Errors (Response Not Recorded) (Voice-to-Text)	Errors in Responses (Manual)	Average Time per MCQ Paper (Voice-to-Text) (mins)	Average Time per MCQ Paper Manual (secs)	Evaluator (1 or 2)
1	200	2.23	12.34	2	1	0.617	37.02	1
2	200	3.91	12.67	1	3	0.6335	38.01	1
3	200	3.45	14.23	0	1	0.7115	42.69	1
4	200	4.11	12.89	0	0	0.6445	38.67	1
5	200	2.67	11.52	1	2	0.576	34.56	1
6	200	2.78	14.13	1	1	0.7065	42.39	2
7	200	2.33	10.89	1	0	0.5445	32.67	2
8	200	2.9	11.73	1	0	0.5865	35.19	2
9	200	3.12	13.37	0	1	0.6685	40.11	2
10	200	2.88	15.1	0	3	0.755	45.3	2
11	200	3.34	13.5	0	1	0.675	40.5	2
12	200	3.11	13.75	1	2	0.6875	41.25	2
13	200	2.95	12.21	0	1	0.6105	36.63	2
14	200	3.01	12.75	2	1	0.6375	38.25	2
15	200	2.42	15.1	2	0	0.755	45.3	2
Total	**3000**	**45.9**	**194.82**	**12**	**13**	-	-	-

Discussion

Error Analysis

The evaluation process highlighted distinct types of errors between the manual and automated (Voice-to-Text and Excel automation) methods. In the manual method, errors were often due to human oversight, such as marking correct responses as incorrect or vice versa, even when other responses in the sequence were accurately marked (**Figure 7**) and **Table 3**. This aligns with findings from previous studies that emphasize human error in manual grading due to fatigue or distraction (Haladyna, Downing, & Rodriguez, 2002). In contrast, errors in the Voice-to-Text (VTT) method were primarily technical, such as incomplete recordings where only 9 responses were captured instead of 10, as depicted in **Figure 6** and **Table 3**.

Issues with Google Docs or LLM occasionally necessitated refreshing the page or dealing with internet connectivity problems, which is consistent with challenges noted in the use of automated systems (Peat, 2000). Additionally, there were instances where LLM provided outputs in formats other than the requested table, requiring further corrections. **Figure 4** illustrates the distribution of errors between the VTT and manual methods, highlighting that while the types of errors differ, both methods contributed to the overall error count. These technical errors highlight the potential for improving the reliability of automated systems with better calibration and error-checking protocols.

Interpretation of Findings

The use of multiple-choice questions (MCQs) is a standard practice in theory examinations across all subjects in medical colleges (Schuwirth & van der Vleuten, 2011). However, many institutions face limitations such as insufficient staffing and lack of advanced grading technologies like Optical Mark Recognition (OMR) sheets and scanners. This is particularly challenging when the number of MCQ sheets per subject ranges from 200 to 500, as shown in **Table 1** and **Figure 5**, which illustrate the extensive time and error rates associated with manual grading methods.. The automated Voice-to-Text and Excel method offers a practical and cost-effective alternative, requiring minimal infrastructure, making it highly accessible even in resource-constrained settings. This approach also allows for the digital sharing of marked responses, enhancing transparency and providing a valuable resource for post-examination review. The adoption of such technology can alleviate some of the burdens faced by educational institutions with limited resources, improving both accuracy and efficiency (Case & Swanson, 2002). **Figure 3** and **Table 2** clearly demonstrate the significant reduction in time and evaluator fatigue achieved with the VTT method compared to manual grading, further supporting its use in educational settings.

Comparison with Previous Studies

A review of the literature indicates limited research on using voice-to-text and Excel automation for MCQ evaluation in medical education. Current technologies like Optical Character Recognition (OCR) are widely used for grading MCQs but necessitate OMR sheets and specialized scanners, which are not feasible in all settings (Brown et al., 1999). The Voice-to-Text and Excel method, by contrast, does not require additional materials or equipment beyond a standard computer and internet access. This simplicity makes it a versatile tool suitable for diverse educational environments, especially those with limited resources. **Figure 2** provides a visual representation of the error-checking setup in Excel for the VTT method, illustrating its straightforward implementation compared to more complex OCR systems. The absence of studies directly comparing these methods underscores the innovation of this approach, suggesting a promising area for further research and development in educational assessments.

By addressing both the logistical and technical challenges associated with traditional MCQ grading methods, this study demonstrates the potential of integrating accessible technologies to streamline the assessment process and reduce human error, contributing significantly to the field of medical education.

Challenges

This study faced a few minor challenges. A quiet environment was preferred for optimal dictation accuracy, but this can easily be arranged in most settings. The free version of LLM has response limits, but upgrading to a paid subscription can easily overcome this for larger datasets. We did not explore variations in results based on factors like evaluator experience, though this is unlikely to significantly impact the overall findings. While evaluators were aware of being timed, which might have introduced a slight bias, their natural pace can be maintained with minimal influence. Technical challenges, such as occasional dictation errors and Excel's limitations, are manageable and do not significantly hinder the broader application of these methods.

Conclusion

Integrating voice-to-text technology and Excel automation in MCQ evaluation significantly improves efficiency and maintains high accuracy. This approach offers the potential for enhanced error detection and correction without the limitations of human fatigue, making it a valuable tool for medical education. The ability to streamline the evaluation process while reducing errors provides a reliable and scalable solution that can adapt to various educational environments.

Future Directions

Future research could explore the integration of scanning technology to automate the grading of handwritten or printed MCQ papers, using AI tools like LLM to evaluate image-based assessments or convert scanned papers into a structured table format. Additionally, expanding this method to other types of assessments beyond MCQs could be valuable, given its simplicity and improved efficacy. Further studies could investigate the use of different automation tools and technologies to enhance the versatility and applicability of this approach across various educational settings and subjects.

Conflict of Interest Statement: The authors declare no conflicts of interest related to this study.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
2. Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2011). General overview of the theories used in assessment: AMEE Guide No. 57. *Medical Teacher*, 33(10), 783-797.
3. Case, S. M., & Swanson, D. B. (2002). Constructing written test questions for the basic and clinical sciences. National Board of Medical Examiners.
4. Peat, M. (2000). Online assessment: The use of web-based self-assessment materials to support self-directed learning. *Assessment & Evaluation in Higher Education*, 25(4), 347-356.
5. Tobin, K. G. (1998). The use of concept mapping to reveal prospective science teachers' prior knowledge. *International Journal of Science Education*, 10(5), 491-508.
6. Brown, G., Bull, J., & Pendlebury, M. (1999). Assessing student learning in higher education. Routledge.